



32nd Annual **INCOSE**
international symposium
hybrid event
Detroit, MI, USA
June 25 - 30, 2022

Evaluating the Effectiveness of a Digital Engineering Tool in a Project Engineering Company

Thomas Saraby Vatlé
University of South-Eastern Norway
Hasbergsvei 36
3616 Kongsberg, Norway
thomas.saraby.vatle@gmail.com

YangYang Zhao
University of Oslo
Postboks 1080
0316 Oslo, Norway
yangyang.zhao@sfe.uio.no

Kristin Falk
University of South-Eastern Norway
Hasbergsvei 36
3616 Kongsberg, Norway
kristin.falk@usn.no

Copyright © 2022 by Thomas Saraby Vatlé, YangYang Zhao, and Kristin Falk. Permission granted to INCOSE to publish and use.

Abstract. The research for this paper focused on the evaluation of digital engineering tools that aid industrial engineering processes. The paper aims to investigate how to evaluate the effectiveness of such a tool to provide insight into how it fits the needs of a project engineering company. The research carried out a case study in a Norwegian subsidiary of a worldwide consulting and project engineering company. The company had previously implemented the digital engineering tool SmartPlant for information management, engineering drawings, interdisciplinary collaboration, and data exchange. The focus of the industry case study was to customize and use an evaluation method in practice on SmartPlant. The research used an evaluation method based on quality characteristics from the ISO/IEC 25010:2011 standard. We collected qualitative data on the digital engineering tools' usage in practice from semi-structured interviews and a survey. In addition to this, we investigated the appropriate analysis methods for survey results, such as mean value vs. net promoter score. The research concluded that the evaluation method provided the company with valuable insight into the tool's benefits and drawbacks.

Introduction

Company. The case company (hereby called: company or case company) is a project engineering company servicing the Norwegian oil & gas industry, and is part of a larger international organization. The Norwegian organization of the company specializes in topside project execution (oil platforms and ships/vessels).

Industry project. The industry project in which we carry out the research is an engineering, procurement, construction & installation project (EPCI). It involves a subsea tie-in to an existing client/customer (hereby called: client) owned oil platform in the North Sea. The subsea tie-in is a project to develop and connect a petroleum discovery to existing infrastructure on the Norwegian continental shelf. The company is the contractor for the "Topside EPCI," the detailed engineering project to facilitate the tie-in of new wells to the platform. The project timeframe is four-year.

The Topside EPCI project includes (as of 02.2021) 111 project personnel from the company in Norway and internationally. Since the focus of the research is on evaluating digital engineering tools, the needs of the technical engineering disciplines using these are in focus. The research includes the following engineering disciplines: process, instrument & automation, electrical, piping, mechanical and heating, ventilation & air-conditioning (HVAC).

SmartPlant. The case study evaluates the digital engineering tool SmartPlant, software made by Hexagon AB. The case company uses the tool in the Topside EPCI. The Hexagon product catalog states that SmartPlant intends to facilitate “seamless collaboration across the value chain, engineering, information management, reduction of errors, data handover and client communication” (Hexagon AB 2019). SmartPlant helps engineering and managing processing plants (incl. oil platforms) and has a “smart” database for equipment information (such as valves, instruments, pipes, electrical equipment). One of its purposes in the case company is to improve the handling of lifecycle information (LCI). LCI is a term used in the Norwegian oil & gas industry to denominate a large set of documentation and data supplied by contractors to client companies during project lifecycles (Fidjeland 2014). Other potential benefits are improved information flow and collaboration between the engineering disciplines and 2D drawing functionality with equipment database integration (smart drawings). The tool has built-in software for making drawings such as piping & instrumentation diagrams (P&ID), system control diagrams (SCD), and 3D models.

Challenges. Oil platforms are large infrastructural entities, “systems of systems.” Modifications on offshore platforms involve numerous systems, subsystems, areas, and equipment. These projects involve extensive data, specifications, and interfaces between domains/disciplines to manage during the lifecycle. Managing client deliverables is also crucial. Another challenge is managing multi-year projects involving several cooperating companies and a revolving door of personnel. Handling the engineering process is therefore critical for projects to run smoothly. A company can manage challenges like these, in part by adopting digital engineering tools.

The company has an identified need to improve the engineering process, including handling LCI, drawings/diagrams, equipment information, and data exchange between disciplines and client deliverables. The company acquired SmartPlant in 2007 to improve on these aspects. Preliminary stakeholder discussions revealed a concern that the tool might not be as beneficial as expected.

Evaluating the digital engineering tool can provide valuable information on whether or not it is worth persevering with SmartPlant. However, there is a lack of methods to aid the company in evaluating and identifying benefits and drawbacks in the current usage.

Research question. To address the challenge, we aim to solve the following research question: How to evaluate the effectiveness of a digital engineering tool like SmartPlant that facilitates engineering processes in a medium-sized project engineering company?

The rationale for solving the research question is twofold. The research value comes from identifying an evaluation method for digital engineering tools and testing this in practice on SmartPlant. The industry-specific research value for the company comes from identifying pros, cons and making recommendations for improvements/actions.

Literature review

“Digital engineering tools” is a broad term of software that aids engineers and companies in their engineering processes. This study focuses on evaluating the effectiveness of SmartPlant in the case company. Different tools have different capabilities and serve different needs in different companies. Selecting the right tool for the job is therefore crucial. We aim to investigate the evaluation of such tools and identify and test a method of assessing whether a digital engineering tool fulfills the company’s needs of facilitating an effective engineering process. We review existing literature to

identify the state of the art, focusing on three main aspects: “what to evaluate,” “how to evaluate,” and “for what purpose, in terms of achieving an effective engineering process.”

Evaluation of digital engineering tools. Viewing the digital engineering tool as an instrument to realize company goals, we can use the “enterprise evaluation & assessment” definition of evaluation from the systems engineering body of knowledge (Martin et al., 2020). It defines the purpose of evaluation as “assessing whether an enterprise moves in the right direction by collecting data to measure whether their strategy is working as intended and to identify risks, opportunities, problems and actions” (Martin et al. 2020). This research focuses on whether a “digital engineering tool” works as intended instead of a “strategy.”

There exists proven knowledge of software evaluation in the systems engineering body of knowledge. “ISO/IEC 25010:2011: Systems and Software Engineering – Systems and Software Quality Requirements and Evaluation (SQuaRE)” is a resource for an evaluation framework for software used in engineering. The standard includes a set of generalized evaluation criteria in the form of quality characteristics. There is also a plethora of other literature on how to evaluate software. Jackson et al. (2011) presented a criteria-based assessment of software quality with a human factors point-of-view. Another paper discussed the evaluation of software engineering and tools (Kitchenham et al. 1997), where they carried out a project with the intent of developing and validating an evaluation method. Kitchenham et al. (1997) presented two types of evaluation methods, quantitative and qualitative. They further recommended qualitative methods for cases where the goal was to establish how well a tool fit the organization’s needs and was based on a feature-based evaluation/survey by people who used the software in advance of the evaluation. This qualitative approach is applicable in our case, where there is an existing user base of SmartPlant in the company.

However, to our knowledge, there is a lack of research specifically on evaluating the effectiveness of digital engineering tools in the vein of SmartPlant. One integral part of SmartPlant, the “smart drawing” functionality, has had some previous research. A Bachelor’s thesis, “The Application of Smart Piping and Instrumentation Diagrams” (Kassas et al., 2007), explored the benefits of using “smart P&ID’s” in a small biodiesel pilot project. It found that smart P&ID’s had little benefits to the efficiency of their small-scale project but suggested that larger projects would reap increased benefits from implementing tools like these. It also discussed critical aspects, such as their tool having a steep learning curve and the need to customize the software for the needs of the company/process.

Effective engineering process. To answer our research question, it is essential first to understand what denominates an effective engineering process. The goal of introducing changes, such as new software into a company, is usually to streamline or otherwise improve on certain aspects related to the performance of the engineering processes. There are four main goals of making changes to an existing engineering process: cost reduction, time reduction, improve output quality and improve quality of work-life (Davenport et al., 1990). Thus, “effective engineering process” can be summarized as facilitating engineering better systems with reduced time and budget and satisfied engineers. These processes can potentially benefit from digital engineering tools.

Evaluation criteria

To evaluate the overall effectiveness of a digital engineering tool, we adopt the quality characteristics from the ISO/IEC 25010 standard as evaluation criteria. According to the standard, the two different categories of software performance characteristics are “quality in use” and “product quality model.” The quality in use characteristics in ISO/IEC 25010 encompass “the degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use.” Quality in use includes the human factors aspects of software evaluation and how systems/software interact with their users. The category consists of five characteristics, as shown in Table 1. The product quality model from ISO/IEC 25010 defines criteria that “relate to static and dynamic properties of software.” The cat-

egory consists of eight characteristics, shown in Table 1. Our case study selects characteristics/criteria from these based on a stakeholder and needs analysis.

Table 1. Quality characteristics from the ISO/IEC25010:2011 standard.

Quality in use characteristics	Product quality model characteristics
User effectiveness	Functional suitability
User efficiency	Performance efficiency
User satisfaction	Compatibility
Freedom from risk	Usability
Context coverage	Reliability
	Security
	Maintainability
	Portability

We can relate Davenport et al.'s (1990) theory to ISO/IEC 25010 by using the product quality model to evaluate the digital engineering tools' ability to deliver improved output quality and the quality in use model to evaluate the tools' ability to improve the quality of work-life. To satisfactorily measure and evaluate cost and time reduction requires quantitative data, which, as we will discuss, is difficult in our case.

Systems engineering methods

Systems engineering framework. We carry out our research loosely following the “systems engineering framework” lifecycle model (Sols 2014). From Sols' framework, we incorporate the specific activities of problem formulation, stakeholder identification, stakeholder needs & requirements, selection of design concept (the evaluation method in our case), testing (using the evaluation method), validation, and verification. Using systems engineering methods allows for structuring the approach and ensuring research validation according to stakeholder needs.

To evaluate SmartPlant as a digital engineering tool, we identify and involve relevant stakeholders. These form the basis for our stakeholder needs, requirements, and validation (Faisandier et al. 2020). We use stakeholder needs to derive stakeholder requirements used to select and validate the evaluation criteria.

Validation in systems engineering is the process of judging a system or solution against the customer's needs, meaning that the final system is according to what the client wanted in the first place (Sols 2014). We involve stakeholders throughout the research process for validation purposes with interviews and surveys, from problem identification to assessing the results in the end.

Verification is the process of confirming how well a system/solution fulfills a set of requirements (Snoderly et al., 2020). In our case, we use verification to confirm the selected evaluation criteria with stakeholders and assess how well survey results fulfill company acceptance criteria. We elaborate on the latter in the research methods section. Verifying the research outcomes is not feasible in the scope of this paper; instead, the focus is on verifying the evaluation criteria and survey results according to stakeholder needs.

Iterative approach. We used an iterative approach on aspects in this research. An iterative approach is often used in software and systems engineering and facilitates agile development through an incremental or evolutionary process (Forsberg 2020). In the case of survey design, we make use of an iterative and evolutionary process. We iterated the survey design in three steps. The first resulted in a set of preliminary questions based solely on stakeholder needs. The second iteration produced an updated set of questions after the literature review and our solidified research methods. The final iteration resulted in the refined survey questions grouped and aligned with ISO/IEC 25010 quality

characteristics to allow for evaluation. The last survey design iteration is the one we distribute for data collection purposes and is the focal point of the “case study” section.

Research methods

The study presented in this thesis uses a systems engineering research method with a set of research activities shown in Figure 1. The research divides into two main parts, an industry case study, and a literature review, the latter shown in green in the figure. The primary outcome of the literature review is the identification of an evaluation framework, which we adapt to our specific industry case. The main research activities shown in blue in Figure 1 include problem identification, stakeholder needs, analysis, evaluation, validation & verification. Red indicates specific means of data collection used in the research, semi-structured interviews, review of company documentation, and a survey.

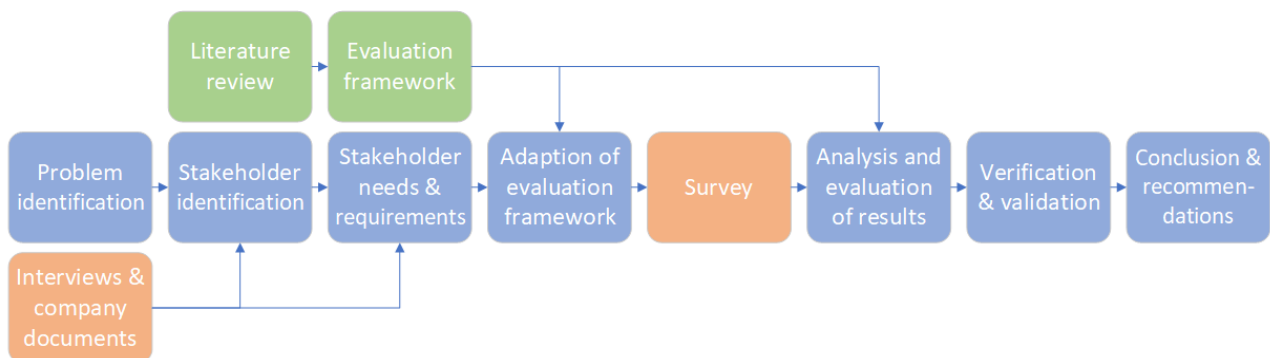


Figure 1. The flow of research methods used in this study. Blue shows research activities, green shows literature review, and red shows specific means of data collection.

Qualitative research. We carry out the case study as indicated in blue and red in Figure 1. We decided that a qualitative research approach was the most appropriate in this case, based on suggestions in (Kitchenham et al. 1997). Another rationale for this approach is a lack of quantifiable data on time and money saved using SmartPlant in the company. We base the research on aspects discussed in (Szajnfarder et al. 2017). This paper gave examples of how to structure qualitative research and essential aspects to keep in mind. These aspects include the importance of selecting the right focus group and designing appropriate interviews and surveys in the case study. An additional note is on one of the goals of effective engineering processes, “time reduction” (Davenport et al., 1990). We can include data on perceived time reduction by SmartPlant users to replace verifying this with additional quantitative time measurements.

Collection of data. Secondhand data is data and documentation that already exists in the company. Secondhand data in our case includes historical use of SmartPlant in the company, documentation on the intended usage of the tool by the different engineering disciplines, its implementation, and interfaces. Firsthand data is collected actively by the author using three main methods: semi-structured interviews, survey, and practical usage of SmartPlant. The survey intends to collect a quantitative dataset from qualitative opinions from SmartPlant users in the case Topside EPCI project to uncover the as-is situation. The survey result is the basis on which we evaluate the tool.

Semi-structured interviews. To uncover the needs for digital engineering tools, we use a series of semi-structured interviews and discussions with relevant stakeholders. We include “main principal lead” engineers, SmartPlant system administrators and specialists, and a subset of project engineers and lead engineers. We use the stakeholder interviews primarily for needs identification as input to survey design and evaluation and for research validation and verification.

Likert scale survey. We use a survey to measure how SmartPlant scores among its company userbase according to the respective evaluation criteria/characteristics. To get quantifiable measures from a qualitative survey, we use a 5-point Likert scale. Rensis Likert introduced his scale in (Likert

1932), which collected data on survey respondents' attitudes towards a question, statement, or similar. The scale included the options "strongly agree," "agree," "neutral," "disagree," and "strongly disagree." We use methods for "Likert scale" as opposed to "Likert-type" (Boone, Jr. et al. 2012), where evaluation criteria were scored based on a composite result of multiple survey questions.

Selection of survey participants. The identified personnel to include in the SmartPlant survey are the engineers that have had at least some experience with SmartPlant in the project. We include the engineering disciplines: process, instrument & automation, electrical, piping, mechanical & HVAC.

Likert scale mean. Boone, Jr. et al. (2012) suggested using the mean value when analyzing Likert scale results, also when using composite scores from multiple questions to evaluate. We will calculate the mean Likert scale scores for each of the selected evaluation criteria. First, we calculate the mean value for each question. Each Likert scale response produces a numerical value: strongly agree = 5, agree = 4, neutral = 3, disagree = 2 & strongly disagree = 1. The result is a mean value from 5 to 1 for each question. If multiple questions are required to answer a single evaluation criterion, we take the mean of the values calculated from each question in the category. The result is the composite mean score. We use the uniform 5-point scale suggested by Pimentel (Pimentel 2019) to draw meaning from the scores, shown in Table 2. For verification purposes, we combine the scale with a set of company acceptance criteria in the case study.

Table 2. Uniform 5-point scale for evaluation (Pimentel 2019).

Score range	5 - 4,2	4,19 - 3,4	3,39 - 2,6	2,59 - 1,8	1,79 - 1
Likert scale	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Net promoter score. In addition to the mean values, we use the net promoter score (NPS) as an additional measure of the effectiveness of SmartPlant. Reichheld described NPS in (p. 7 in Reichheld et. al. 2011). His NPS used a 10-point scale of satisfaction from a collected survey dataset. Answers 10 and 9 indicated respondents regarded as promoters, 8-7 indicated passives, and 6-1 were detractors. The calculated NPS was then the percentage of promoters minus the percentage of detractors. The defining characteristic of the NPS was that it only regarded the most positive answers as promoters, indicating the respondents who benefit from a solution/product the most, whereas 8-7 meant that they were satisfied but not remarkably so and therefore regarded as passives.

We use the 5-point scale analog to the NPS on our Likert-scale results, where "strongly agree" represents promoters, "agree" indicates passives, and "neutral, disagree, and strongly disagree" are the detractors. The new scale ranges from 100% to -100%. 100% indicates that every non-neutral respondent is a "promoter," and vice versa. Note that the main difference from the mean calculation is that this method regards "agree" as a neutral response and "neutral" as negative.

Case study

We carry out the case study focusing on evaluating the digital engineering tool SmartPlant and its effectiveness in the Topside EPCI project. This section mainly elaborates on the case analysis, case findings, validation, and verification.

Stakeholder and needs analysis in the Topside EPCI

We uncovered relevant stakeholders and their needs from company documentation and semi-structured interviews. The company had documentation explaining the intended SmartPlant responsibilities for the technical disciplines, aiding in the initial stakeholder identification. We uncovered their needs through interviews.

Direct stakeholders. The direct stakeholders are those who use SmartPlant directly in day-to-day tasks. These stakeholders include project engineers using SmartPlant, discipline lead engineers,

“tools” personnel (those responsible for implementing and supporting digital tools in the company), SmartPlant system administrators, and international company personnel using SmartPlant.

Indirect stakeholders. We also considered the needs of the indirect stakeholders, who have stakes in SmartPlant but do not use the tool actively. These include the client company, department lead engineers, document control, project management, company management, and line organization.

Stakeholder needs. The result from the stakeholder analysis is a set of stakeholder needs. The needs for digital engineering tools we uncovered are: intuitive in use, a low threshold of use, facilitate effective exchange of data, facilitate the production of standard deliverables & drawings, ability to handle LCI, facilitate effective change management, facilitate effective interdisciplinary and international cooperation, save time, prevent user error, reliability, and satisfied users.

Evaluation criteria

We use the stakeholder needs to select the relevant assessment criteria based on the quality characteristics from ISO/IEC 25010. The quality characteristics unrelated to our stakeholder needs, we eliminate from the evaluation. To evaluate the overall effectiveness of SmartPlant as a digital engineering tool, we include criteria from both the quality in use characteristics and the product quality model. Table 3 shows the traceability between stakeholder needs and the eight selected relevant criteria with sub-characteristics.

Table 3. The selected evaluation criteria from ISO/IEC 25010 related to stakeholder needs.

Stakeholder needs	ISO/IEC 25010:2011		
	Quality in use	Product quality	Sub characteristic
LCI/Information management, exchange of data, production of standard deliverables		Functional suitability	Functional completeness
Production of standard deliverables, change management, interdisciplinary communication			Functional appropriateness
Save time, prevent user error		Usability	User error protection
Intuitive, Low threshold of usage			Learnability
Intuitive, Low threshold of usage			Operability
Save time		Performance Efficiency	Time behavior
Reliability		Reliability	Availability
Exchange of data, LCI/information management		Compatibility	Interoperability
LCI/Information management, exchange of data, save time, international cooperation	User effectiveness		
Save time	User efficiency		
Satisfied users	User satisfaction		

Functional suitability. Functional suitability is among the most critical evaluation criteria in our case. The company implemented SmartPlant to serve a set of functions, particularly LCI/information management, smart drawing capabilities with database integration, and change management. If the tool does not effectively serve its functional purpose, it will not be effective overall. We included the sub-characteristics in Table 3 to show the need for the tool to have the necessary functionality (functional completeness) and appropriate functionality, i.e., that the functions work well.

Usability. There is a strong need in the company for the tool to be usable. In our case, this means the intrinsic usability of the tool should include being learnable, operable, and protect from user error. Learnability entails the ease with which new users can start using the tool. Operability is the need for the tool to be intuitive in use. User error protection is another relevant sub-characteristic, as effective computer-aided design helps reduce simple human mistakes and ultimately save time on rework.

Performance efficiency. We also evaluate the perceived performance efficiency of SmartPlant by its users. The performance efficiency means the “time-behavior” of the tool itself, whether it has fast

response times and is not lagging or excessively time-consuming to use.

Reliability. We also evaluate the reliability based on experiences from SmartPlant users. In our case, this entails the tool's availability; it should be available for use when needed and not crash.

Compatibility. Compatibility includes the capabilities for the tool to exchange data with other interfacing software/systems. In our case, SmartPlant has an interface to the client company database for LCI transfer, and we survey whether this interface is perceived to be effective/compatible.

User effectiveness. We evaluate the user effectiveness based on how well SmartPlant helps the users achieve the tool-specific goals.

User efficiency. The user efficiency relates to the "resource expenditure" needed to achieve the user goals of using SmartPlant. Here the term resource expenditure means the amount of time spent/saved and if it helps reduce the number of unnecessary tasks.

User satisfaction. We finally evaluate user satisfaction, which we use as a general measure of the positivity to SmartPlant by its users.

Survey design. We use the list of stakeholder needs and the selected evaluation criteria in Table 3 to design survey questions. Most of the criteria require answering several survey questions to capture the necessary aspects.

Especially in two particular cases, we need multiple questions to gather data on the relevant aspects, functional suitability, and usability. Regarding functional suitability, we include questions related to the main desired functionalities we gathered from the stakeholder interviews: information management, drawings, database, database integration, and change management. In usability, we ask questions related to operability, learnability, and user error protection. We investigate individual questions more specifically when receiving varying results within a single category in the survey results and discussion section.

Survey results and discussion

In this section, we introduce and discuss the survey results. The survey initially had 28 potential respondents and received 19 responses, resulting in a response rate of 68%. The discussion combines data gathered through stakeholder interviews, discussions, and the survey with author reflections. To analyze the results from multiple viewpoints, we divide them into categories: overall (including all 19 respondents), engineering discipline, years of SmartPlant experience, and industry experience. We begin by discussing the mean results before comparing these with the NPS results.

Mean results. We calculate the composite mean scores from the survey results for each of the eight evaluation criteria. The scale adapted from Table 2 includes a set of stakeholder acceptance criteria for verification purposes, shown in Table 4. We use this scale to evaluate our mean results.

Table 4. Scale for evaluation of mean results, including color-coding and stakeholder acceptance criteria "satisfactory," "acceptable," "improvements needed," and "not acceptable."

Score range	5 - 4,2	4,1 - 3,4	3,3 - 2,6	2,5 - 1,8	1,7 - 1
Likert scale	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Acceptance criteria	Satisfactory	Acceptable	Improvements needed	Not acceptable	

Table 5 shows the overall composite mean scores. It shows an acceptable score in total and in functional suitability, performance efficiency, reliability, user effectiveness, efficiency, and satisfaction. We derive from this is that SmartPlant functionality overall seems to fit the company's needs but has room for improvement. The quality in use characteristics all score in the acceptable range, pointing

towards the users being positive towards the tool. Most categories score acceptably, but we note that out of these, they score in the lower end of the acceptable range, from 3,6 to 3,4.

There are two categories with “neutral/improvements needed” in the overall scores: usability and compatibility. The respondents seem to perceive SmartPlant usability as low, with the lowest score being only 3,1. Stakeholders proposed usability as among the most critical aspects. The low score indicates that it is challenging to learn and operate SmartPlant.

When diving by technical discipline in Table 5, there is an observed variation in results. There is a higher satisfaction among the process engineers than the other disciplines, including the only two criteria in the “strongly agree/satisfactory” range. These are functional suitability and user effectiveness, both with a score of 4,2. This result indicates that SmartPlant functionality fits the process discipline needs. Combining this with findings from our interviews, we can attribute this in part to the process discipline having utilized more of SmartPlant capabilities for a longer time and having a larger userbase. This usage includes being the first discipline in the company to create drawings in SmartPlant, mainly P&ID’s. The total score for the process discipline is 4,0, which is the highest in the dataset.

There is also a significant area of concern when investigating the results divided by discipline. All the non-process disciplines end up with a total score in or close to the “neutral/improvements needed” category. According to our acceptance criteria, this implies the need to consider measures to improve the scores. We receive predominantly negative scores from the electrical and instrument & automation disciplines. From instrument & automation, we have the two lowest scores in the “not acceptable” range, with usability at 2,4 and performance efficiency at only 2,0. The interviews uncovered a specific area of concern of the software being slow and cumbersome in use, particularly as a drawing tool. This opinion could explain the low score in performance efficiency.

Table 5. Color-coded composite mean score survey results. The results are divided by the overall score from all 19 respondents and by engineering discipline. Dark green indicates a score in the “strongly agree” range, light green “agree,” yellow “neutral,” and orange “disagree.”

ISO/IEC 25010:2011		Overall Composite Mean Scores	Composite Mean Score by Engineering Discipline			
			Electrical	Instrument & Automation	Process	Piping, Mechanical & HVAC
Quality in Use	Product Quality	19 responses	5 responses	3 responses	7 responses	4 responses
	Functional Suitability	3,5	2,7	3,1	4,2	3,1
	Usability	3,1	2,9	2,4	3,7	2,7
	Performance Efficiency	3,4	2,8	2,0	4,1	4,0
	Reliability	3,5	3,0	3,0	4,1	3,5
	Compatibility	3,3	2,8	3,3	3,7	3,0
User Effectiveness		3,6	3,0	3,0	4,2	3,5
User efficiency		3,4	3,0	2,7	4,0	3,3
User Satisfaction		3,6	3,2	2,6	4,1	3,8
Total:		3,4	2,9	2,8	4,0	3,4

One of the most significant results emerges when dividing by years of SmartPlant experience in Table 6. The results show a significant upwards trend in the total scores correlating with increased experience from 3,2 in those with less than one year’s experience to 3,9 in those with more than three years of experience. In this case, compatibility and user satisfaction score at 4,2, in the “strongly agree/satisfactory” range. It is worth noting that we observe a slight dip in the score from “<1 year” to “1-3 years” experience. Variations due to the low sample size of three in the latter category could cause this. From these results, we can assume that the increased perceived effectiveness of SmartPlant correlates with increased experience. From this, we gather that it might be worth sticking

with the tool, even if new users struggle.

We also investigate the results when dividing by years of industry experience in Table 6. Here we receive the most favorable scores from those with between 10 and 20 years of experience in the industry. The results indicate that it seems to be easier to see the benefits of SmartPlant once an individual has more work experience. It is also likely that these respondents have experience from previous projects that did not use SmartPlant, thus having more perspective with which to compare. Finally, we see primarily unsatisfactory results in those with 0-10 years of industry experience in Table 6. The least favorable result is again the performance efficiency category at 2,5. The result could imply that younger engineers do not readily accept slow response times in the software.

Table 6. Color-coded composite mean score survey results. The table depicts results when dividing by years of experience with SmartPlant (SP) or the industry in general. Dark green indicates a score in the “strongly agree” range, green “agree,” yellow “neutral,” and orange “disagree.”

ISO/IEC 25010:2011		Composite Mean Score by SP Experience			Composite Mean Score by Industry Experience		
		<1 year	1-3 years	>3years	0-10 years	10-20 years	>20 years
Quality in Use	Product Quality	10 responses	3 responses	6 responses	4 responses	9 responses	6 responses
	Functional Suitability	3,4	2,8	3,9	3,0	3,7	3,2
	Usability	2,9	2,9	3,4	2,8	3,4	2,7
	Performance Efficiency	3,0	3,0	4,0	2,5	3,7	3,7
	Reliability	3,4	3,0	3,8	3,3	3,8	3,3
	Compatibility	2,9	3,0	4,2	3,3	3,4	3,3
User Effectiveness		3,4	3,0	3,9	3,0	3,8	3,7
User Efficiency		3,1	3,2	3,8	3,1	3,4	3,6
User Satisfaction		3,2	3,5	4,2	3,3	3,6	3,8
Total:		3,2	3,0	3,9	3,0	3,6	3,4

Functional suitability tradeoff. We find another interesting variance in the results when looking into the specific survey questions relating to functional suitability. The three specific functions surveyed were LCI/information management, drawing tool, and change management (see **Error! Reference source not found.** in the appendix). We receive the following mean values: 4,2 (LCI/information management), 3,6 (change management) and 3,2 (drawing functionality). The results indicate more positivity towards SmartPlant as an information management/LCI-tool than as a drawing tool with change management coming in-between. From survey results and stakeholder discussions, SmartPlant was perceived not to be as effective as a drawing tool as other software with which the respondents had previous experience.

We can further divide the “drawing functionality” results into the “creation of drawings” and “drawing/database integration.” Then the results show: 2,8 (drawing functionality) and 4,1 (drawing/database integration). On the one hand, this shows an overall negative view regarding the tool’s effectiveness for drawings, but a score close to satisfactory regarding integration between drawings and database (smart drawings). Suppose we chose to accept that SmartPlant is not as effective a drawing tool as other software. In that case, we could presume that it still has significant benefits stemming from the equipment database integration. We found the same opinion during interviews: “it is cumbersome to create drawings in SmartPlant, but the database integration is still a large benefit if used appropriately.”

Net promoter score results. To test the credibility of using the mean value to evaluate the results, we analyze the results again using calculated NPS percentages. We can then compare these scores to the mean values to look for similar trends or differences. We employ five uniform intervals for evaluation in in Table 7, with acceptance criteria following the scale used for the mean results in

Table 4. To analyze results, we divide them into the same acceptance criteria as for the mean results.

Our interviews uncovered concerns of a certain amount of negativity and resistance towards SmartPlant from project personnel in the past. The NPS scores could help us gain a better understanding of why this is the case. In the literature review, we found that the theory behind the NPS (Reichheld et al. 2011) suggested that only the most positive respondents would be “fighting” for SmartPlant to be kept and further incorporated. The scale also suggests that those who respond with “agree” are generally neutral towards whether the company keeps the tool or not.

Table 7. Scale used for evaluation of NPS including color coding and corresponding stakeholder acceptance criteria “satisfactory,” “acceptable,” “improvements needed,” and “not acceptable.”

Score range	100% to 60%	59% to 20%	19% to -20%	-21% to -60%	-61% to -100%
Acceptance criteria	Satisfactory	Acceptable	Improvements needed	Not acceptable	

Looking at Table 8 **Error! Not a valid bookmark self-reference.**, we immediately notice the stark contrast to the mean results in Table 5. Every evaluation criteria in the overall composite NPS results score as “not acceptable.” The net percentage of detractors ranges between -23% in user satisfaction to -49% in usability.

Most of the results in our NPS analysis show similar trends to the mean results but with a lower baseline. We will not go into the same amount of detail as in the discussion of our mean results, but note the similarities, contrasts, and implications of these. Table 8 shows again that the process discipline is the most positive towards the tool. Although, this time, it is only coming in with a total of 6% net promoters. We also note that this is the only individual category that yields a net score above zero. From our stakeholder acceptance criteria, this still constitutes “improvements needed.” It is not desirable that there is only a slightly higher amount of promoters compared to detractors when the goal is for the tool to be effective and have a satisfied user base in the company. This time the electrical discipline scores the lowest. On average, there is a -83% net percentage of electrical engineers opposed to SmartPlant compared to promoters, evidently not acceptable.

Table 8. Color-coded composite NPS percentage survey results. The results are divided by the overall score from all 19 respondents and by engineering discipline. Yellow indicates a score in the “improvements needed” range, orange and red as “not acceptable.”

ISO/IEC 25010:2011		Overall Composite NPS %	Composite NPS % by Engineering Discipline			
			Electrical	Instrument & Automation	Process	Piping, Mechanical & HVAC
Quality in Use	Product Quality	19 responses	5 responses	3 responses	7 responses	4 responses
	Functional Suitability	-25	-77	-48	17	-36
	Usability	-49	-78	-79	-11	-67
	Performance Efficiency	-38	-100	-100	14	0
	Reliability	-35	-80	-67	14	-50
	Compatibility	-40	-100	0	-14	-100
User Effectiveness		-32	-87	-56	19	-33
User Efficiency		-41	-80	-58	-4	-46
User Satisfaction		-23	-60	-58	14	-17
Total:		-35	-83	-58	6	-44

Table 9 shows the results divided by years of SmartPlant experience and industry experience. This time the trends are primarily the same as for the mean results but again with a lower baseline. The number of promoters increases with gained SmartPlant experience and shows that those with between 10 and 20 years of experience are less negative. For those with less than one years’ experience with SmartPlant, we again receive the lowest score in usability at -67% net amount of detractors.

These results seem to indicate that the training of new users is unsuccessful. The category of engineers with three or more years of SmartPlant experience is the only division where we receive acceptable scores with the NPS method, with compatibility and user satisfaction both coming in with a net 20% of promoters. This score is at the lower end of the acceptable range and should not indicate that the tools' effectiveness is satisfactory. The engineers with the fewest years of industry experience again score the lowest. Here, every criterion score in the lowest category between -75% in reliability, compatibility, and user satisfaction, to the bottom of the scale, with performance efficiency coming in at -100%.

Table 9. Color-coded composite NPS percentage survey results. The table depicts results when dividing by years of experience with SmartPlant (SP) or the industry in general. Light green indicates a score in "acceptable" range, yellow in "improvements needed," orange and red as "not acceptable".

ISO/IEC 25010:2011		Composite NPS % by SP Experience			Composite NPS % by Industry Experience		
		<1 year	1-3 years	>3years	0-10 years	10-20 years	>20 years
Quality in Use	Product Quality	10 responses	3 responses	6 responses	4 responses	9 responses	6 responses
	Functional Suitability	-35	-52	0	-77	-2	-42
	Usability	-67	-46	-29	-88	-22	-68
	Performance Efficiency	-63	-50	0	-100	-11	-33
	Reliability	-44	-50	-17	-75	0	-75
	Compatibility	-63	-100	20	-75	-25	-33
User Effectiveness		-44	-78	0	-92	-7	-27
User Efficiency		-58	-54	-8	-88	-28	-28
User Satisfaction		-51	-17	20	-75	-6	-23
Total:		-53	-56	-2	-84	-13	-41

We can deduct from the NPS results that this could be why company management has observed negativity towards SmartPlant. A fair amount of "agree" Likert responses drag up the mean results showing a more favorable situation, but this is not the complete picture. When assuming that "agree" respondents are indifferent to SmartPlant being persisted with or discarded, we see that the detractors outnumber the promoters. Judging by this, it might be worth considering alternatives to SmartPlant.

Research validation and verification

This section presents validation and verification of the research methods. We discussed the usefulness of the research results with stakeholders, whether we surveyed the right personnel and asked the right questions. We also reflect on how different analysis methods could lead to different results/conclusions.

An important question to consider during validation is whether we reached the industry-specific goals of the research. This goal includes identifying the effectiveness of SmartPlant, its benefits and drawbacks, and whether it is worth it for the company to continue or abandon it. We have validated the application of the evaluation method. By applying the methods from ISO/IEC 25010, we could offer insights into the specific characteristics determining how effective a digital engineering tool is. We uncovered specific areas of concern that need improvement to be effective, particularly usability, performance efficiency, compatibility, and specific functions like the drawing functionality. There are also specific subsets of engineers who are likely to benefit from improved training in the tool's usage. By dividing the results into subsets of respondents, we identified more specifically where improvements are needed. The areas we uncovered where SmartPlant works well suggest that it could be worth implementing the tool into projects. These areas include its effectiveness for LCI management and the effectiveness for engineers seems to correlate with increased experience.

Regarding verification of the research methods, we are concerned with how well the applied methods helped to evaluate the effectiveness of the digital engineering tool. To verify the selection of evaluation criteria, we discussed these with stakeholders. The question is whether we eliminated ISO/IEC 25010 criteria that the evaluation should have included. The eliminated criteria were: freedom from risk, context coverage, security, maintainability, and portability. The first two criteria we agreed with stakeholders were not relevant in our case due to the fixed context of use (office/home office) and there not being hazardous risks involved. We also eliminated the last three criteria due to evaluating based on a userbase survey, where the general users do not have insights into these aspects.

Given the survey evaluation results, the initial stakeholder reactions to the mean survey results were favorable. The results came as a positive surprise, challenging stakeholder suspicions that SmartPlant would score more poorly overall. The result's rigorousness depends on whether we included the right personnel and asked the right questions in the data collection. We collected data from the SmartPlant userbase in the Topside EPCI. While we included personnel from the most relevant disciplines, there is a question of whether this was enough. The survey response rate was only at 68%, which resulted in an underrepresentation of some respondent groups. Increasing the sample size could increase the results' reliability. We agreed with the company on the potential benefits if we had iterated on the survey design once more before distributing it. Another iteration could have further improved the alignment with ISO/IEC 25010 quality characteristics.

To further verify whether the results came from research method limitations, we introduced a second analysis method. With the NPS method, the results showed a different situation, more aligned with management expectations. The different analysis methods mean value and NPS produced different results in our case. The key takeaway from our research is that to evaluate a digital engineering tool effectively, there are benefits to analyzing the results from multiple viewpoints. If we had only evaluated the overall mean survey results, we could have concluded that six out of eight characteristics scored acceptably, and only two needed improvements. However, this does not reflect the entire situation. Using NPS revealed a situation where all the overall quality characteristics scored as "not acceptable." The differences illustrate that selecting proper analysis methods for Likert survey results is paramount. We also reflect on the possibilities that other methods than a survey could have produced different results, like in-depth userbase interviews. A survey "forces" responses into sets of predefined questions and alternatives, which depth interviews would not (Ogden et al., 2012).

Conclusion

We have discussed the benefits of understanding stakeholder needs to effectively evaluate digital engineering tools, done on a case-by-case basis. Our research contribution is to showcase how to apply systems engineering methods and customize the existing ISO/IEC 25010 standard quality characteristics to evaluate a digital engineering tool in a project engineering company. We combined stakeholder needs and quality characteristics with a userbase survey to measure the effectiveness of SmartPlant in the case company. From this, we found both benefits and areas that need improvement. Finally, we conclude on the research question. One approach to effectively evaluate digital engineering tools is to survey the tool's userbase by combining ISO/IEC 25010 quality characteristics with stakeholder needs, investigating the results from subsets of respondents, and combining analysis methods like mean value and NPS to gain a more holistic understanding.

Recommendations

The industry-specific goal of the case study was to provide insights into the effectiveness of SmartPlant in the project engineering and make recommendations for improvements. We made a set of recommendations from our reflections on the stakeholder discussions and survey results.

The first recommendation is to introduce SmartPlant to more project personnel; we observe a rise in effectiveness as the engineers gain more experience with SmartPlant and become familiar with its

functionalities. The company will likely get more personnel over the initial barrier/threshold usage by introducing more engineers to the tool quicker. A second recommendation is to consider more extensive training in using the tool. We generally receive poor usability scores, which indicates that the training has not been satisfactory. This aspect is especially relevant for the non-process disciplines from whom we received poor results. Increasing the number of SmartPlant users in the engineering disciplines with a smaller userbase could also aid in the training aspect by increased experience transfer between users. The third recommendation is to investigate technical measures to improve SmartPlant reliability (reduce software crashes) and responsiveness (performance efficiency). The fourth recommendation is to carry out another evaluation by the end of the Topside EPCI project to see if the results have improved or not. Finally, we recommend investigating possibilities of introducing SmartPlant into more projects, not only of the detailed engineering variety.

We also provide general recommendations not specific to the case company. The first recommendation is that it is beneficial for companies to properly investigate and do a preliminary evaluation of a digital engineering tool before spending years adopting, implementing, and adapting it. The final recommendation is to carry statistics on time usage and money spent vs. saved to help verify results from effectiveness evaluations.

Research limitations and future research

We researched one digital engineering tool in one company in one project during the early phase of a single lifecycle stage. In our case, we researched the effectiveness of SmartPlant in the early phase of the detail engineering stage of the Topside EPCI project in the case company. This implies that we cannot conclude our methods as successful in a generic sense. What we have done is to demonstrate one evaluation method to be useful for one digital engineering tool. By carrying out the research early in the Topside EPCI project, it limited the ability to see changes or improvements over time. Many project members were also still new to the case software.

The main research limitation was a lack of quantitative data on time usage and money spent vs. saved by implementing SmartPlant and its reliability. Therefore, we could not fully evaluate the aspects of cost reduction and time reduction from Davenport's paper (Davenport et al., 1990). Instead, we evaluated time and reliability based on opinions from userbase experiences and did not verify based on quantified data. The most important aspect not investigated is the monetary aspect due to a lack of data. Answering whether there are cost benefits requires further research.

Another research limitation emerges from using a Likert scale survey as the primary means to gather quantitative data from qualitative opinions. Previous studies discussed limitations when using Likert scale data to make decisions: (Ogden et al. 2012) and (Theofanidis et al. 2018). A limitation to keep in mind is differences in data when comparing Likert scale results with "free text accounts" (interviews/conversations in our case) (Ogden et al., 2012). Another limitation is how respondents interpret survey questions and response alternatives differently (Ogden et al. 2012). Finally, there is also a tendency for people not to select the extreme points but land somewhere in the middle even if they are positive or negative (Theofanidis et al., 2018).

A final noteworthy research limitation is the low sample size in some of our categories. An example is in Table 5, where we divided the results by technical discipline. In this case, we only received three responses from instrument & automation engineers. The low sample size means that individual opinion significantly impacts the final score than with greater sample size. A final limitation worth noting is that due to time constraints, we could not investigate whether there are long-term benefits that have not materialized yet.

There is a need for future research to supplement and verify our research methods and results. Our research only evaluated one digital engineering tool in a particular industry case. We propose using the same methods in other case studies to investigate their effectiveness in evaluating different digital

engineering tools in different companies. A future research proposal specific to the case company is to repeat the survey close to the end of the Topside EPCI to investigate whether the results have improved or not.

References

- Boone Jr., H. N., and D. A. Boone 2012, '*Analyzing Likert Data*', Journal of Extension Volume 50 Number 2, West Virginia University (USA).
- Davenport, T. H., and J. E. Short 1990, '*The New Industrial Engineering: Information Technology and Business Process Redesign*', Center for Information Systems Research, Massachusetts Institute of Technology (USA).
- Faisandier, A., Roedler, G., and R. Adcock 2020, '*Stakeholder Needs and Requirements*', SEBoK v. 2.3, viewed 15.04.2021, <https://www.sebokwiki.org/wiki/Stakeholder_Needs_and_Requirements>.
- Fidjeland, B. 2014, '*LCI – Document package or continuous process?*', infuseit.com, viewed 12.03.2021, <<http://www.infuseit.com/Main-menu/Blog/SubPages/2014/SubPages/LCI--Document-package-or-continuous-process/>>
- Forsberg, K. 2020, '*System Life Cycle Process Models: Iterative*', SEBoK v. 2.3, viewed 15.04.2021, <https://www.sebokwiki.org/wiki/System_Life_Cycle_Process_Models:_Iterative>.
- Hexagon AB 2019, '*SmartPlant Foundation: Visualizing Key Information From Complex Data Sources*', Product Sheet, Hexagon AB (Sweden).
- ISO/IEC 25010:2011(E), '*Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuARE) – System and software quality models*', International Standard.
- Jackson, M., Crouch, S., and R. Baxter 2011, '*Software Evaluation: Criteria-based Assessment*', Software Sustainability Institute, University of Edinburgh (UK).
- Kassas, A., and A. Gikas 2017, '*The Application of Smart Piping and Instrumentation Diagrams*', Worcester Polytechnic Institute, Massachusetts (USA).
- Kitchenham, B., Linkman, S., and D. Law 1997, '*DESMET: a methodology for evaluating software engineering methods and tools*', Computing & Control Engineering Journal.
- Likert, R. 1932, '*A Technique for the Measurement of Attitudes*', Archives of Psychology, 22 (140), 1-55.
- Martin, J., Lawson, B., and A. Faisandier 2020, '*Enterprise Systems Engineering Process Activities*', SEBoK v. 2.3, viewed 08.05.2021, <https://www.sebokwiki.org/wiki/Enterprise_Systems_Engineering_Process_Activities>.
- Ogden, J., and J. Lo 2012, '*How Meaningful are data from Likert scales?: an evaluation of how ratings are made and the role of the response shift in the socially disadvantaged*', International Journal of Health Psychology, 12, 350-361.
- Pimentel, J. 2019, '*Some Biases in Likert Scaling Usage and its Correction*', International Journal of Sciences: Basic and Applied Research ISSN 2307-4531.
- Reichheld, F., and R. Markey 2011, '*The Ultimate Question 2.0, How Net Promoter Companies Thrive in a Customer-Driven World*', Bain & Company, Harvard Business Review Press, Boston (USA).
- Snoderly, J., and A. Faisandier 2020, '*System Verification*', SEBoK v. 2.3, viewed 16.05.2021, <https://www.sebokwiki.org/wiki/System_Verification>.
- Sols, A. 2014, '*Systems Engineering Theory and Practice*', Universidad Pontificia Comillas, Madrid (Spain).
- Szajnfarder, Z., and E. Gralla 2017, '*Qualitative Methods for Engineering Systems: why we need them and how to use them*', The George Washington University, Washington DC (USA).
- Theofanidis, D., and A. Fountouki 2019, '*Limitations and Delimitation in the Research Process*', Perioperative nursing (GORNA), E-ISSN:2241-3634, 7(3), 155–162.